DOI:10.11931/guihaia.gxzw201907051

# 阴地蕨全转录组分析及植物激素信号转导相关基因筛选

张林甦，韩忠耀，王传明，邓先扩
（黔南民族医学高等专科学校 药学系，贵州 都匀 558000）

**摘要：**阴地蕨是蕨类阴地蕨科阴地蕨属常用植物药，其生长发育有一定的代表性，为获得其转录组等生物学信息进行了二代测序和分析。该文以新鲜阴地蕨全株为材料，用 Illumina HiSeq 2500 平台进行全转录组测序，干净序列经组装后得单一基因（Unigene），将 Unigene 在非冗余蛋白/核酸数据库（nonredundant protein database,Nr）、核酸序列数据库（Nucleotide Sequence Database，Nt）、基因本体论数据库（gene ontology,GO）、蛋白质真核同源数据库（clusters of eukaryotic orthologous groups,COG）、东京基因与基因组百科全书（Kyoto encyclopedia of genes and genomes,KEGG）、蛋白质序列数据库 SwissProt 和 Interpro 进行生物信息学分析。结果显示共获得干净序列 6.67 Gb ，组装后得到 58 646 个 Unigene，平均长度 1 023 bp，Unigene 在上述数据库的总体注释率为 69.25%，其中，在 GO 数据库中 20 762 个基因被注释到生物功能、细胞组分和分子功能 3 个本体的 52 个功能组，COG 注释了 20 633 个基因并将其划分为 25 个功能簇，在 KEGG 数据库中比对注释了 29 377 个基因，可划分为 5 个大类、19 个亚类代谢途径,据此筛选出八类植物激素信号转导相关的 41 个基因家族；通过比对共得到 43 102 个编码序列（coding sequence，CDS），平均长度 749 bp, N50 为 1 137；筛选到 60 个转录因子 transcript factor,TF）家族共 1 502 个转录因子基因；共发现 17 195 个单核苷酸多态性（single-nucleotide polymorphism，SNP）位点，其中碱基转换 11 122 个，颠换 6 073 个；发现了 8 245 个简单序列重复（simple sequence repeat，SSR），数量最多的为二核苷酸重复和三核苷酸重复。这些结果从功能和结构方面提供了阴地蕨整体转录组信息和涉及植物激素信号转导的潜在基因，为进一步深入研究阴地蕨的生长发育、遗传、品种鉴别等提供了分子生物学的基础数据。

**关键词：**阴地蕨，转录组，植物激素，信号转导，基因筛选



# Global transcriptome analysis of *Botrychium ternatum* and screening of its plant hormone signal transduction related genes

ZHAND Linsu， HAN Zhongyao， WANG Chuanming , DENG Xiankuo
( *Department of Pharmacy, Qiannan Medical College for Nationalities*，
Duyun 558000, Guizhou, China )

**基金项目：**本研究获得贵州科技基金(QKH-JC-[2016]1137), 黔南科技基金(QNKHSZ[2016]-2)；黔南民族医学高等专科学校基金(QNYZ-201602、QNYZ-201801)支持[Supported by Guizhou Science and Technology Foundation (QKH-JC-[2016]1137); Qianna Science and Technology Foundation (QNKHSZ[2016]-2); Qiannan Medical College for Nationalities Foundation (QNYZ-201602, QNYZ-201801)]。

**作者简介：**张林甦（1972－），女，贵州贵阳人，博士，副教授，主要从事药用植物功能基因组学研究，(E-mail) linsuzhang009@163.com。

**Abstract:**
*Botrychium ternatum* is a commonly used folk medicinal plant, its growth and development have shown some typical representative characters of some fern plants . But researches on it mainly focus on chemical constituents, clinical and pharmaco-logical effects, classification and distribution investigation, few about its molecular biology.Plant hormone is a sort of small signal molecular and has very important function to plant growth and development, and plant hormone signal transduction play key role in hormonal equilibrium. To obtain related information, Illumina HiSeq 2500 platform was used to perform transcriptome sequencing and bioinformatics analysis were carried out afterwards. Results showed 6.67 Gb clean reads was obtained and 58 646 Unigenes were assembled with an average length of 1 023 bp. Unigenes were annotated in NR, NT, GO, COG, KEGG,Swissprot and Interpro databases with an overall annotation rate of 69.25%.Through GO annotation, 20 762 genes were annotated to three terms and 52 functional groups. 20 633 genes were divided into 25 functional cluster by COG annotation. Through KEGG analysis, 29 377 genes were mapped to 5 groups and 19 sub-groups pathways. In addition, 41 gene families related to eight plant hormone signal transduction were screened. Through BLAST and ESTScan, 43 102 coding sequences (CDS) were found, with average length 749 bp, N50 1 137. 60 Transcript Factor gene families with total 1 520 genes were screened out, including C3H, MYB,MYB-related, bHLH,AP2-EREBP,WRKY and GRAS. 17 195 single-nucleotide polymorphisms (SNP) were found, inclding 11 122 transitions and 6 073 transversion. And 8 245 Simple sequence repeats were found, among them di-nucleotide repeats and tri-nucleotide repeats rank the top two abundance. Those data sets provide valuable source for further research on the growth and development of *B ternatum*.

**Key words:** *Botrychium ternatum*, transcriptome, plant hormone, signal transduction, gene screening



植物激素是一类信号小分子，对于植物的生长发育起着重要作用，其通过植物激素信号转导系统发挥作用：内因或外因可促使一系列植物激素基因诱导和表达，作用于相应的激素受体或组件，最终显现出不同的性状（苏谦等，2008）。常见的植物激素有：生长素、细胞分裂素、赤霉素、脱落酸、乙烯、油菜内酯素、茉莉酸和水杨酸等，在这些激素信号转导系统中有的受体或关键组分因互作（interact）或串话（crosstalk）会产生协同或拮抗的作用而使信号途径网络化（Ohri et al., 2015）。比如光信号可以通过与生长素信号途径发生串话从而调控根的发育（Kumari & Panigrahi, 2019）,光敏色素作用因子（PIFs）可以响应赤霉素、油菜内酯素、茉莉酸、生长素吲哚乙酸(IAA)、脱落酸、乙烯等信号途径，通过这个“枢纽”分子将激素信号途径联成了复杂的网络（任小芸等，2016）。植物激素还可通过表观遗传调控促进开花，如赤霉素、茉莉酸、脱落酸和生长素对 DNA 甲基化、组蛋白翻译后修饰介导的染色质压缩起重要作用从而影响开花（Campos-Rivero et al., 2017）。另外，在植物应对土壤病原菌侵害时也进化出复杂的激素信号网络来保护自身（Berens et al., 2017）。可见植物的激素信号转导系统对植物的生长发育、防御、环境适应等具有重要意义。

阴地蕨（*Botrychium ternatum*），又名一朵云、小春花、蛇不见、郎其细辛、独脚蒿、冬草等，属于阴地蕨科阴地蕨属，一年生草本药用植物，多以孢子进行繁殖，其生长发育具有一定的代表性。是我国民间，尤其是贵州、福建等地常用一种的中

草药，因其具有清热解毒、止咳、止血等功效，主要用于小儿高热惊搐，肺热咳嗽，咳血，百日咳，毒蛇咬伤，目赤火眼，目生翳障等（齐建红，2012；赵俊华等，2008；阮君山，2002）。目前对阴地蕨的研究较少，主要集中在化学成分、临床及药理作用、分类及分布调查等方面，分子生物学相关的信息较少，限制了更深入的研究。转录组（transcriptome）是指某一生理条件下，细胞内所有转录产物的集合，包括信使 RNA（mRNA）、核糖体 RNA（rRNA）、转运 RNA（tRNA）及非编码 RNA(none coding RNA)。随着测序技术的发展和普及，转录组测序（RNA-seq）已经成为从分子水平研究生物基因及其调控的重要方法。本研究通过高通量测序获得阴地蕨全转录组，通过生物信息学方法对其进行分析，得到阴地蕨转录组的整体注释信息、筛选出植物激素信号转导相关的潜在基因及其单核苷酸多态性（single nucleotide polymorphism, SNP）和短序列重复多态性（short sequence repeat polymorphism, SSR）等信息，为进一步从分子水平开展阴地蕨生长发育、品种鉴定等研究提供了有用的资源。

## 1 材料与方法

### 1.1 材料

新鲜、成熟阴地蕨植物全株 3 株（包含根、茎、叶及孢子，经黔南医学高等专科学校王传明副教授鉴定为阴地蕨）2016 年 7 月采于贵州省黔南州都匀市郊斗篷山地区（海拔约 1 500 m，东经 107°20′-107°27′E ，北纬 26°12′-26°16′N），立即用清水冲洗干净、吸水纸吸干后放入干冰盒中带回，备提 RNA。

### 1.2 cDNA 文库制备及测序

将植物全株用液氮研磨成粉末，用 RNA 提取试剂盒（艾德莱公司，北京）提取总 RNA 并将 DNA 消化，用带有寡聚脱氧胸腺嘧啶(Oligo dT)的磁珠富集 mRNA，经琼脂糖电泳及微量核酸检测仪 NanoDrop 检测合格后用试剂盒依次合成 cDNA、纯化、修复粘性末端、在 3’末端加上碱基“A”并连接接头，然后进行片段大小选择，最后进行 PCR 扩增构建 cDNA 文库；构建好的文库经检验合格后上 Illumina HiSeq 2500 平台进行测序。

### 1.3 De novo（从头） 组装

将测序得到的原始序列（raw reads）去除低质量、接头污染以及未知碱基 N 含量过高的序列得到干净序列（clean reads），使用 Trinity 软件（v2.0.6）（Grabherr et al.,2011 ）对 clean reads 进行 de novo 组装，然后使用 Tgicl 软件（v2.0.6）（Pertea et al., 2003）将组装的转录本进行聚类去冗余，得到单一基因（Unigene）用于后续分析。

### 1.4 Unigene 功能注释及分析

为了解 unigene 的功能，用生信分析软件将 unigene 在七大功能数据库中进行注释：用 Blast (v2.2.23) 进行 NT、NR、COG、KEGG 以及 SwissProt 注释，用 Blast2GO (v2.5.0)(Conesa et al., 2005) 以及 NR 注释结果进行 GO 注释，用 InterProScan5 (v5.11-51.0) (Quevillon , et al., 2005)进行 InterPro 注释。根据 KEGG 信号途径 map04075，将经注释的相关基因进行归类，即得植物激素信号转导相关基因。

### 1.5 转录组结构分析

CDS（编码序列，Coding sequences）预测： 根据功能注释结果，按照 NR，SwissProt，KEGG，COG 的数据库优先顺序，挑选 Unigene 的最佳比对片段作为该 Unigene 的 CDS 。未能注释上的 Unigene 使用预测得到的 CDS 作为模型进行建模，然后使用 ESTScan (v3.0.2) ( Iseli et al., 1999)进行 CDS 预测。

TF（转录因子，Transcript factor）编码能力预测：用 getorf(EMBOSS:6.5.7.0) ( Rice et al., 2000) 检测 Unigene 的开放阅读框（ORF，Open reading frame），使用 hmmsearch(v3.0) ( Mistry et al., 2013) 将 ORF 比对到转录因子蛋白结构域(数据来源

于 PlantTFDB)，然后根据 PlantTFDB 描述的转录因子家族特征对 Unigene 进行 TF 编码能力鉴定(Jin et al., 2017)。

SSR 和 SNP 检测：用 MISA （v1.0）(Thiel et al., 2003) 对 Unigene 进行 SSR 检测；用 HISAT（ v0.1.6-beta）(Kim et al., 2015) 把 clean reads 比对到 Unigene，然后使用 GATK （v3.4-0）(McKenna A et al., 2010) 检测 SNP。

## 2 结果与分析

### 2.1 测序及组装结果

本项目使用 Illumina Hiseq 平台一共测得总原始序列（raw reads）数据量为 55.52 Mb,过滤后得到干净序列（clean reads）44.45 Mb, clean reads 比率达到 80.6%，测序深度属“深度”（high deep，>15 Mb）测序。得到干净总碱基数 6.67 Gb ，组装后得到 58 646 个 Unigene，平均长度 1 023 bp，N50 、N70 均大于 1 000 bp (表 1) 。所有 Unigene 的长度均大于 300 bp, 分布在 300-400 bp 的最多, 占 25.5%，大于 1 000 bp 的累计占 39%（图 1），说明测序连续性和组装效果较好。

表 1 Unigene 的质量指标

Table 1 Quality metrics of Unigene assemble

| 样本<br>Sample | 总数<br>Total number | 平均长度、<br>Mean length | N50 | N70 | N90 | GC(%) |
|---|---|---|---|---|---|---|
| 阴地蕨<br>*Botrychium_ternatum* | 58 646 | 1 023 | 1 762 | 1 102 | 399 | 44.64 |

注：**N50.** 按 Unigene 长度从大到小排序后逐个累加至所有 Unigene，总长度的 50%时，最后一个累加的数值大小即为 N50；N70、N90 以此类推。**GC(%).** 碱基 G 和 C 的比例。

Note：**N50.** A weighted median statistic that 50% of the total length is contained in Unigenes great than or equal to this value. The same of N70 and N90. **GC(%).** Percentage of G and C bases in all Unigenes.



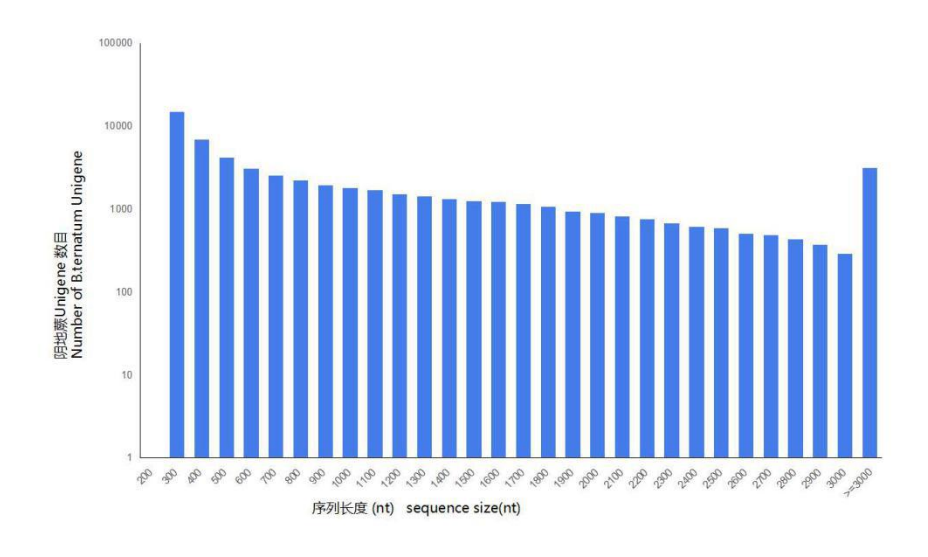



图 1

*Botrychium ternatum* Unigene 的长度分布

Fig. 1 Length distribution of *Botrychium ternatum* Unigene

2.2 Unigene 功能注释

将 Unigene 进行七大功能数据库注释(NR, NT, GO, COG, KEGG,Swissprot 和 Interpro)，注释结果见表 2。在 NR( NCBI 蛋白数据库，NCBI protein database)中得到最多注释（65.4%），总体注释率为 69.25%。根据 NR 注释结果统计了注释物种分布（见图 2），在蕨类植物小立碗藓（*Physcomitrella patens*）和江南卷柏（*Selaginella moellendorffii*）中共注释了 24%，跟阴地蕨蕨类植物属性相符，另外在常用的参比物种北美云杉（*Picea sitchensis*）中注释也较高（14.21%），可能跟北美云杉本身的注释较好有关（Ralph et al., 2008）。NR、COG(直系同源基因簇数据库，Clusters of Orthologous Groups database)、KEGG(京都基因及基因组数据库，Kyoto Encyclopedia of Genes and Genomes database)、Swissprot(Swiss 蛋白数据库) 以及 Interpro（Interpro 蛋白数据库）的注释结果展示在图 3，在五个数据库中都注释上的有 12 522 个，占全部 unigene 的 21.4%。



表 2 *B.ternatum* Unigene 功能注释结果统计
Table 2 Statistic of function annotation of *B.ternatum* Unigene

| 数值 Values | Unigene 总数 Unigene Total | Nr 注释 Nr Annotated | Nt 注释 Nt Annotated | Swissprot 注释 Swissprot Annotated | KEGG 注释 KEGG Annotated | COG 注释 COG Annotated | Interpro 注释 Interpro Annotated | GO 注释 GO Annotated | 总体注释 Over all |
|---|---|---|---|---|---|---|---|---|---|
| 数目 Number | 58,646 | 38,352 | 23,391 | 27,049 | 29,241 | 16,678 | 30,375 | 15,905 | 40,610 |
| 占比 Percentage | 100% | 65.40% | 39.89% | 46.12% | 49.86% | 28.44% | 51.79% | 27.12% | 69.25% |

注：总体注释为被七大数据库中任意一个数据库注释上的 Unigene 总数。
Note: Overall is total Unigene annotated by either of the seven databases.

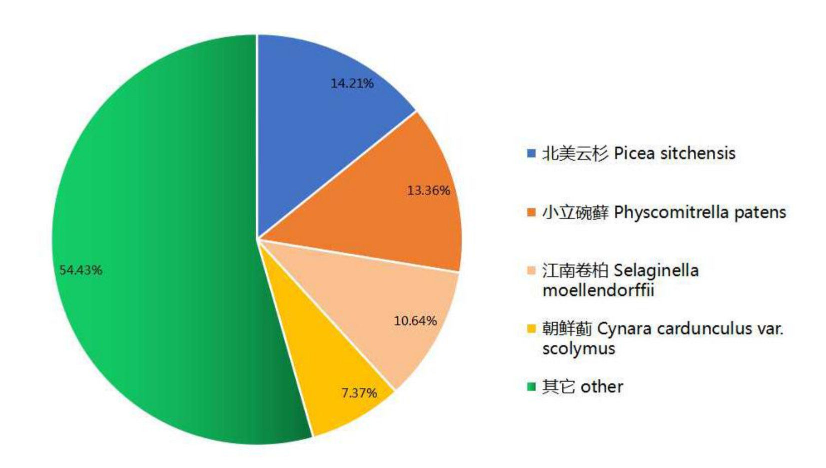



图 2　*B.ternatum* Unigene 注释物种

Fig. 2　Annotation species statistics of *Botrychium ternatum*

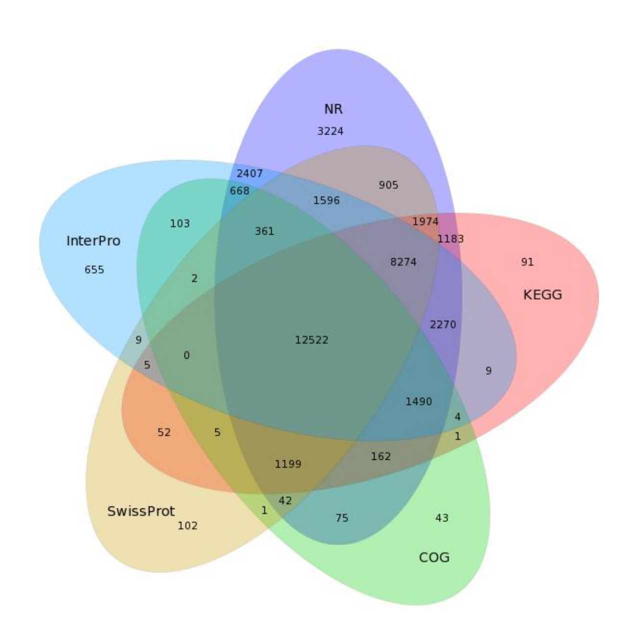
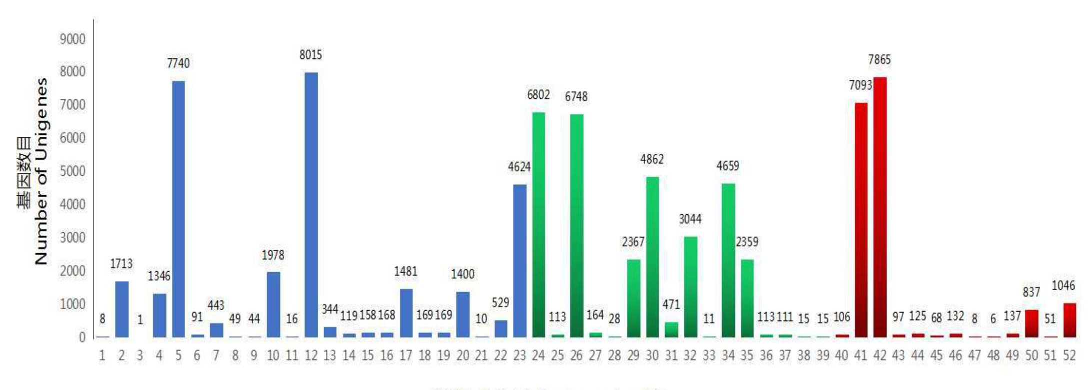



图 3　NR、COG、KEGG、Swissprot 以及 Interpro 的五个数据库注释韦恩图

Fig. 3 Venn diagram between NR, COG, KEGG, Swissprot and Interpro

2.3 GO 注释结果

通过 GO 注释将 20 762 个阴地蕨基因或基因产物赋予三大类术语（term）：分子功能、细胞组分和生物学过程，GO 功能分布如图 3。在生物学过程中涉及基因数量处于前三位的分别是：代谢过程（metabolic process ）、细胞过程(cellular process)，和单组织过程（single-organism process）。细胞组分中最多的是细胞（cell），最少的是核苷（nucleotide）。分子功能中数量最多的是催化活性（catalytic activity）和结合(binding)，其次是转运活性（transport activity）。




注：**1.** 生物粘附；**2.** 生物调节；**3.** 细胞死亡；**4.** 细胞成分组织或生物发生；**5.** 细胞过程 **6.**解毒；**7.**发展过程；**8.**生长；**9.**免疫系统过程；**10.**定位；**11.**运动；**12.**代谢过程；**13.**多组织过程；**14.**多细胞生物的过程；**15.**生物过程的负调控；**16.**生物过程的正向调节；**17.**生物过程调节；**18.**繁殖；**19.**生殖过程；**20.**应激反应；**21.**节律过程；**22.**信号；**23.**单组织的过程；**24.**细胞；**25.**细胞连接；**26.**细胞部分；**27.**细胞外区域；**28.**细胞外区域部分；**29.**大分子复合体；**30.**膜；**31.**膜部分；**32.**膜封闭腔；**33.**类核；**34.**细胞器；**35.**细胞器部分；**36.**超分子纤维；**37.**合胞体；**38.**病毒粒子；**39.**病毒粒子部分；**40.**抗氧化活性；**41.**结合；**42.**催化活性；**43.**电子载体活性；**44.**分子功能调节器；**45.**分子传感器；**46.**核酸结合转录因子活性；**47.**营养储存；**48.**蛋白质标记；**49.**信号传感器；**50.**结构分子；**51.**转录因子活性，蛋白质结合；**52.**运输活性。

**1.** Biological adhesion; **2.** Biological regulation; **3.** Cell killing; **4.** Cellular component organization or biogenesis; **5.** Cellular process; **6.** Detoxification; **7.** Developmental process; **8.** Growth; **9.** Immune system process; **10.** Localization; **11.** Locomotion; **12.** Metabolic process; **13.**Multi-organism process; **14.** Multicellular organismal process; **15.** Negative regulation of biological process; **16.** Positive regulation of biological process; **17.** Regulation of biological process; **18.** Reproduction; **19.** Reproductive process; **20.** Response to stimulus; **21.** Rhythmic process; **22.** Signaling; **23.** Single-organism process; **24.** Cell; **25.** Cell junction; **26.** Cell part; **27.** Extracellular region; **28.** Extracellular region part; **29.** Macromolecular complex; **30.** Membrane; **31.** Membrane part; **32.** Membrane enclosed lumen; **33.** Nucleoid; **34.** Organelle; **35.** Organelle part; **36.** Supramolecular fiber; **37**. Symplast; **38.** virion; **39.** Virion part; **40.** Antioxidant activity; **41.** Binding; **42.** Catalytic activity; **43**. Electron carrier activity; **44.** Molecular function regulator; **45.** Molecular transducer activity; **46.** Nucleic acid binding transcription factor activity; **47**. Nutrient reservoir activity; **48**. Protein tag; **49.** Signal transducer activity; **50**. Structural molecule activity; **51.** Transcription factor activity, protein binding; **52.** Transporter activity.

图 4　GO 功能注释

Fig. 4 GO Function annotation

2.4 COG 功能注释

通过于 COG 数据库进行比对，将 20 633 个阴地蕨 unigene 进行 COG 注释，结果如图 5。聚在一般功能（general function prediction only）的最多（4559 个），包含

1000-2000 个基因的簇有 8 个，包括翻译、核糖体结构及生物发生及转录等重要的生命活动，值得注意的是鉴定了 995 个未知功能（function unknown）基因。

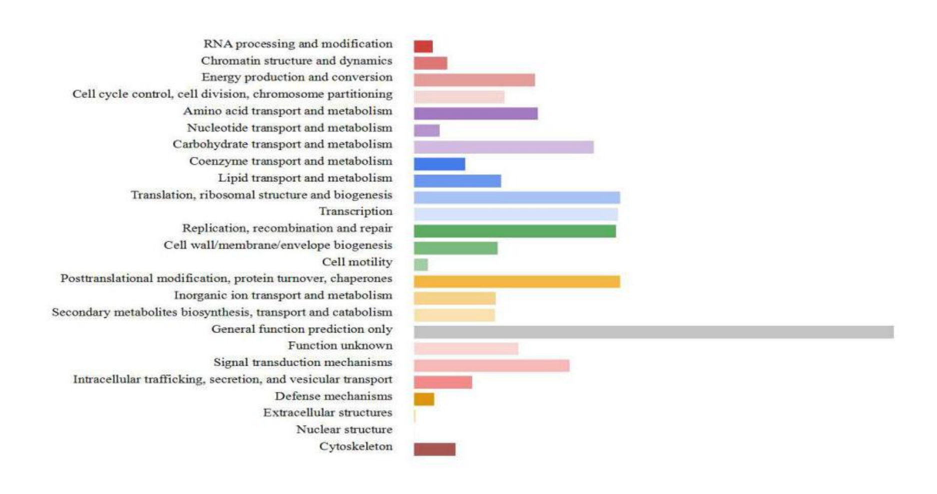



图 5 COG 功能注释
Fig 5 COG Annotation

2.5 KEGG 通路分析及植物激素信号转导基因筛选

共有 29 377 条基因比对到六大类、21 亚类代谢通路上（图 6），其中数量最多的是新陈代谢（metabolism）通路，有 17 698 个基因，占 60%；最少的是与人类疾病相关的基因，共 141 个（因阴地蕨属植物）；与有机系统（organismal system）环境适应（environmental adaption）有关的基因有 1 266 个。根据 KEGG 信号途径 map04075，将经注释的相关基因进行归类，得植物激素信号转导相关的候选基因（表 3）。

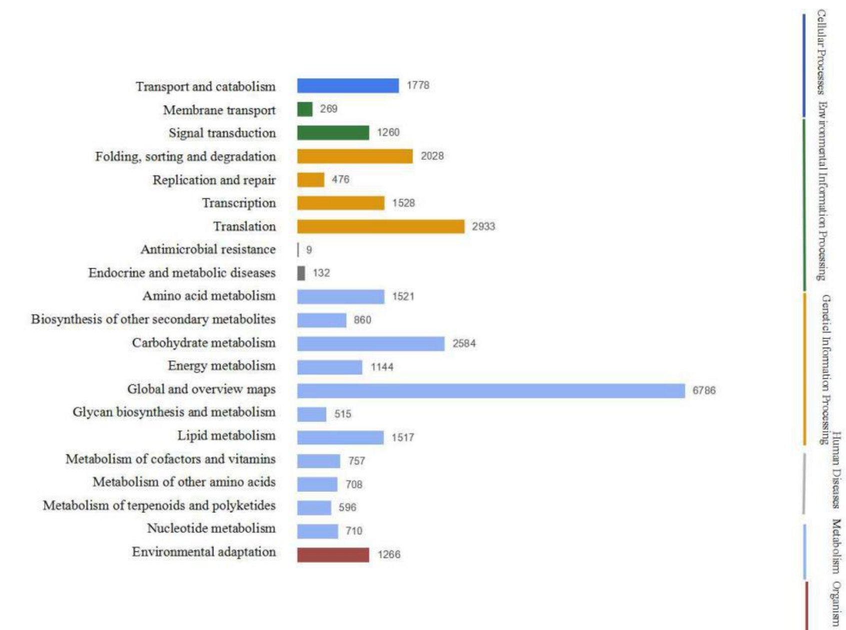



图 6　KEGG 通路分析
Fig. 6 KEGG Pathway Analysis

表 3 阴地蕨植物激素信号转导通路及相关基因
Table 3　Genes in plant hormone signal transduction pathways of *Botrychium ternatum*

| 激素名称<br>Plant hormone | 涉及基因<br>Related genes | 生物作用<br>Biological function |
| --- | --- | --- |
| 生长素 | AUX1(17) | 细胞增大 Cell enlargement |
| Auxin | TIR1(14) | 植物生长 Plant growth |
| | AUX/IAA(23) | |
| | ARF(71) | |
| | GH3(11) | |
| | SAUR(16) | |
| 细胞分裂素 | CRE1(19) | 细胞分裂 Cell division |
| Cytokinine | AHP(8) | 芽形成 Shoot initiation |
| | B-ARR(67) | |
| | A-ARR(7) | |
| 赤霉素 | GID1(18) | 茎生长 Stem growth |
| Gibberellin | GID2(7) | 诱导萌发 Induced germination |
| | DELLA(42) | |
| | TF*(129) | |
| 脱落酸 | PYR/PYL(15) | 气孔关闭 Stomatal closure |
| Abscisic acid | PP2C(48) | 种子休眠 Seed dormancy |
| | SnRK2(23) | |
| | ABF(32) | |
| 油菜素甾醇 | BAK1(39) | 细胞伸长 Cell elongation |
| Brassionsteroid | | |
| | BRI1(83) | 细胞分裂 Cell division |
| | BSK(5) | |
| | BKI1(3) | |
| | BIN2(1) | |
| | BZR1/2(8) | |
| | TCH4(8) | |
| | CYCD3(1) | |
| 乙烯 | ETR(12) | 果实成熟 Fruit ripening |
| Ethylene | CTR1(27) | 老化 Senescence |
| | MPK6(6) | |
| | EIN2(7) | |
| | EBF1/2(6) | |
| | EIN3(13) | |
| | EFR1/2(3) | |
| 茉莉酸 | JAR1(6) | 单萜合成<br>Monoterpenoid biosynthesis |
| Jasmonic acid | COI1(3) | 吲哚生物碱合成<br>Indole alkaloid biosynthesis |
| | JAZ(24) | |
| | MYC2(58) | |
| 水杨酸 | NPR1(3) | 抗病 Disease resistance |
| Salicylic acid | TGA(21) | |
| | PR-1(7) | |

注：括号中数字是基因数目。* 此处 TF 是指 PIF4 和 PIF3。

Note: numbers in brackets are numbers of genes. *TF here indicates PIF4 and PIF3.

2.6 转录组结构

CDS：通过 BLAST 得到 38 212 个 CDS，用 ESTScan 方法得到 4 890 个 CDS，共得到 43 102 个 CDS，平均长度 749 bp, N50 为 1 137。

TF：共筛选到 60 个转录因子基因家族共 1 502 个 TF 基因，数量超过 100 的有 C3H MYB 和 MYB-related 以及 bHLH 转录因子家族，其它较多的还有 AP2-EREBP、WRKY、GRAS 等转录因子。

SNP:共发现 17 195 个 SNP 位点，其中碱基转换 11 122 个，包括 A-G 5 452 个，C-T 5 670 个；颠换 6 073 个，包括 A-C 1 444 个，A-T 1 729 个，C-G 1 418 个，G-T 1 482 个。

SSR:最多的是二核苷酸重复，有 3 666 个，其次是三核苷酸重复，3 439 个，接下来依次是单核苷酸重复（563 个）、六核苷酸重复（260）、四核苷酸重复（169）和五核苷酸重复（148）。

## 3 讨论

阴地蕨是民间常用的药用、非模式植物，主要药用成分为黄酮类和多糖类。通过转录组测序（RNA-seq）测得阴地蕨转录组 clean bases 共 6.67 Gb,各项指标显示测序深度、组装读长均较好。对 unigene 在七大数据库中进行注释，发现注释率最高的是 NR 数据库，因 NR 数据库本身数据量大，包含经过验证的和预测的蛋白，所以注释率虽然高，但是注释结果仍然需要结合其它数据库的注释来考虑。从 NR、COG、KEGG、Swissprot 以及 Interpro 的注释结果韦恩图分析（图 3），可以看出有 12 522 个 unigene 是处于这五个数据库的交集，意为着这部分基因的注释可靠性较高。

GO、COG 和 KEGG 注释是基因注释的重要途径，分别从基因本体（gene oncology）、聚类（cluster）以及通路（pathway）进行描述。在 COG 分析中，发现有 995 个功能未知 unigene，占 4.8%。阴地蕨属厚囊蕨纲（Eusporangiopsida）瓶儿小草目（Ophioglossales）植物，是众多蕨类植物中的一员。蕨类植物在地球上分布广、存在时间长，大部分可以形成孢子，处于低等植物向高等植物过度的位置，生长发育过程有自身特色（Christenhusz & Chase , 2014；张开梅等，2016）。将阴地蕨转录组与其它蕨类如石松目的蛇足石杉全株（Yang et al., 2017）转录组进行比较，发现总体注释率均在 55%-60%，KEGG 注释的途径中涉及基因最多的也是代谢，另外这两种植物中对于环境适应的基因也都占到一定比例，说明蕨类植物对于环境的适应调动了不少基因，所以能较好地适应环境变化，成为现存的一类古老植物。因此研究阴地蕨这类植物的生长发育有一定的意义，尤其是那些筛选出来但是功能未知的基因值得进一步探讨。

植物激素是影响植物生长发育的重要因素，通过 KEGG 分析找出了目前已知的八条植物激素信号转导途径中的基因，大部分基因家族包含的基因数目较少，比较方便下一步的克隆、分析及功能鉴定等研究，但是有的基因家族包含的基因数目较多，如赤霉素途径中的 TF 和 DELLA、油菜素甾醇途径中的 BRI1、细胞分裂素途径中的 B-ARR、脱落酸途径中的 PP2C 等都有好几十、甚至上百个成员，这就需要进一步找出关键基因、缩小研究范围，可以通过对不同组织、器官或者不同处理后进行差异表达基因（Different Expressed Genes,DEGs）分析，这也是本研究的局限。另外，虽同属于蕨类植物，但是有些蕨类生长特性差别很大，阴地蕨生长迅速，每年以孢子方式繁殖，蛇足石杉生长缓慢，孢子萌发需要数年(Guo et al.,2009)。对比蛇足石杉植物激素信号转导途径涉及的基因数量（本实验室数据，存放于

http://bigd.big.ac.cn/gsa,登录号 PRJCA001325），发现有的基因家族组成有很大的不同，数量差别达一倍以上：在阴地蕨的赤霉素信号途径中的鉴定出的 TF（PIF4 和 PIF3）、油菜素内酯信号途径 BAK1、BZR1/2 和 TCH4 家族包含更多基因，而在蛇足石杉的分裂素信号途径中 CRE1 和 A-ARR、油菜素内酯信号途径 BSK、脱落酸信号途径 NPR1 和 TGA 基因家族拥有更多基因，暗示以上基因在这两种植物中可能具有不同的生长调控作用，而这两种植物表现出不同的生长模式的原因很有可能与这些基因家族有关。

除了内源性的激素等因素，外界环境如光照、温度、刺激、土壤等等都会对植物的生长发育产生影响。很多研究致力于找到几种影响因素的关键节点（如光温，害虫和微生物等），多组学和系统生物学不失为研究复杂调控网络的手段之一(Meena et al., 2017;Myburg et al., 2019;Choi, 2019)。

**致谢**

感谢北京基因组研究所（BGI）提供技术支持。感谢郭士川先生提供植物样品采集帮助。

**参考文献**

BERENS ML, BERRY HM, MINE A, et al., 2017. Evolution of Hormone Signaling Networks in Plant Defense[J]. Annu Rev Phytopathol, 55: 401-425.

CAMPOS-RIVERO G,OSORIO-MONTALVO P, SANCHEZ-BORGES R, et al., 2017. Plant hormone signaling in flowering: An epigenetic point of view[J]. J Plant Physiol, 214:16-27.

CHOI HK,2019.Translational genomics and multi-omics integrated approaches as a useful strategy for crop breeding[J]. Genes Genom, 41(2): 133–146.

CHRISTENHUSZ MJ & CHASE MW, 2014. Trends and concept in fern classification [J]. Ann Bot, 113(4): 571-94.

CONESA A, GÖTZ S, GARCÍA-GÓMEZ JM, et al., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research[J]. Bioinformatics, 21(18): 3674–3676.

GUO B,XU L, WEI Y, et al.,2009.Research Progress of *Huperzia serrata*[J]. Zhongguo Zhong Yao Za Zhi,(16): 2018~2022. [郭斌，徐玲玲，尉亚辉，等，2009. 千层塔的研究进展[J]. 中国中药杂志，(16): 2018– 2022.]

GRABHERR MG, HAAS BJ, YASSOUR M, et al., 2011.Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq Data[J]. Nat Biotechnol, 29(7): 644-652.

ISELI C, JONGENEEL CV, BUCHER P, 1999. ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences[J]. Proc Int Conf Intell Syst Mol Biol, 99: 138-148.

JIN JP, TIAN F, YANG DC,et al., 2017. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants[J]. Nucleic Acids Res, 45(D1): D1040-D1045.

KIM D, LANGMEAD B, SALZBERG SL,2015.HISAT: a fast spliced aligner with low memory requirements[J]. Nat Methods,12(4): 357-360.

KUMARI S,PANIGRAHI KCS, 2019, Light and auxin signaling cross-talk programme root development in plants[J]. J Biosci, 44(1): 26.

MCKENNA A, HANNA M, BANKS E, et al., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next generation DNA sequencing data[J]. Genome Res, 20(9): 1297-303

MEENA KK, SORTY AM, BITLA UM, et al., 2017. Abiotic Stress Responses and Microbe-Mediated Mitigation in Plants: The Omics Strategies[J]. Front Plant Sci, 8: 172.

MISTRY J, FINN RD, EDDY SR, et al., 2013. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions[J]. Nucleic Acids Res, 41(12): e121.

MYBURG AA,HUSSEY SG,WANG JP, 2019. Systems and synthetic biology of forest trees: A bioengineering paradigm for woody biomass feedstocks[J]. Front Plant Sci, 10: 775.

OHRI P, BHARDWAJ R, BALI S， et al., 2015. The common molecular players in plant hormone crosstalk and signaling[J]. Curr Protein Pept Sci,16(5): 369-388.

PERTEA G, HUANG X, LIANG F, et al., 2003. TIGR Gene Indices clustering tools (TGICL): A soft ware system for fast clustering of large EST datasets[J]. Bioinformatics, 19(5): 651-652.

QI JH, 2012. A summary of recent studies on *Botrychium* Sw [J]. J Xi'an Univ Arts Sci ( Nat Sci Ed), 15（2）：48-50. [齐建红，2012. 阴地蕨属植物研究进展[J].西安文理学院学报自然科学版, 15（2）：48-50.]

QUEVILLON E, SILVENTOINEN V, PILLAI S , et al., 2005. InterProScan: protein domains identifier[J]. Nucleic Acids Res, 33: W116–120.

RALPH SG, CHUN HJ, KOLOSOVA N, et al., 2008. A conifer genomics resource of 200,000 spruce (Picea spp.) ESTs and 6,464 high-quality, sequence-finished full-length cDNAs for Sitka spruce (Picea sitchensis)[J]. BMC Genomics,9:484. doi: 10.1186/1471-2164-9-484.

REN XY,WU MQ,CHEN JM, et al., 2016. The molecular mechanisms of phytochrome interacting factors (PIFs) in phytohormone signaling transduction [J]. J Plant Physiol, 52 (10): 1466-1473. [任小芸，吴美琴，陈建民，等，2016. 光敏色素作用因子 PIFs 参与植物激素信号转导的分子机制[J]，植物生理学报，52 (10): 1466-1473.]

RICE P, LONGDEN I, BLEASBY A, 2000. EMBOSS: the European Molecular Biology Open Software Suite [J]. Trends Genet,16(6): 276-277.

RUAN JS, 2002. Research progress of *Sceptridium ternatum* and its effective ingredients[J]. J Chin Pharm Univ, 33: 328-329.[阮君山，2002.小春花及其有效成分研究进展[J]. 中国药科大学学报，2002,33: 328-329.]

SU Q, AN D,WANG K, 2008. Phytohormone receptors and induced genes in plants[J]. Plant Physiol Mol Biol, 44(6): 1202-1208.[苏谦，安冬，王库，2008.植物激素的受体和诱导基因[J]. 物生理学通讯，44(6): 1202-1208.]

THIEL T, MICHALEK W, VARSHNEY RK, et al., 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.) [J]. Theor Appl Genet, 106(3): 411-22.

YANG M, YOU W, WU S, et al., 2017. Global transcriptome analysis of Huperzia serrata and identification of critical genes involved in the biosynthesis of huperzine A[J]. BMC Genomics, 18: 245.

ZHANG KM,SHEN Y, LIU Y, et al., 2016. Research progress on development and physio-ecology of fern gametophytes[J]. Guihaia，36( 4) : 419-424. [张开梅，沈羽，刘颖，等，2016. 蕨类植物配子体发育与生理生态研究进展[J]. 广西植物， 36( 4) :

419-424.]
ZHAO JH,ZHAO NW,WANG PS, et al., 2008. Study on the species and distribution of *Adiantum* and *Botrychiam* medicinal plants from Tujia medicine of Guizhou Province origin[J]. J Med Pharm Chin Minor, 5: 44-46. [赵俊华，赵能武，王培善，等，2008. 土家药黔产铁线蕨、阴地蕨科药用植物的种类和分布研究[J]. 中国民族医药杂志，5:44-46.]